

**In the Claims:**

Please cancel claims 9, 11-18, 27-28, and 31-50. Please amend claims 1, 4-8, 10, 19, 22-26, and 29-30. The claims are as follows:

1. (Currently amended) A method for document analysis and retrieval, comprising the following steps of performed in the order recited:

accessing a document taxonomy that comprises M categories such that M is at least 2,  
wherein the document taxonomy is based on a subject matter classification in conjunction with  
a collection of stored documents, wherein each category of the M categories has an associated at  
least one category key, wherein the category keys of all M categories collectively consist of N  
unique category keys sequentially ordered and denoted as CATKEY[1], CATKEY[2], ...,  
CATKEY[N];

transmitting, by a remote host in a first computing system to a web service host in a second computing system, a first portion of a document; and

sequentially transmitting, by the remote host to the web service host, at least one additional portion of the document, wherein the first portion and the at least one additional portion collectively comprise the entire document, wherein the entire document is adapted to be reconstructed and subsequently processed via processing said entire document by the web service host, said processing comprising at least one of:

extracting text from said entire document to configure said text in a text format, if said entire document received by said web service host comprises said text in a non-text

format;

generating a plurality of document keys associated with said text from analysis of said text in said text format, if said entire document received by said web service host comprises said text in said text format, or if said web service host has previously performed said extracting such that said text in said text format is available to said web service host; and

generating a document key vector  $V_{DOC}$  of order  $N$ , wherein said generating  $V_{DOC}$  comprises for  $n = 1, 2, \dots, N$ : determining setting  $V_{DOC}[n]$  equal to 1 if the plurality of document keys comprises a document key equal to  $CATKEY[n]$ , otherwise setting  $V_{DOC}[n]$  equal to 0;

after said generating  $V_{DOC}$ , generating a document weight vector  $W_{DOC}$  of order  $N$ , wherein said generating  $V_{DOC}$  comprises for  $n = 1, 2, \dots, N$ : setting  $W_{DOC}[n]$  equal to a first frequency count raised to a power  $P_1$  greater than 1, wherein the first frequency count consists of a number of appearances, in the document, of the document key associated with  $V_{DOC}[n]$  if  $V_{DOC}[n]$  is equal to 1 or consists of 0 if  $V_{DOC}[n]$  is equal to 0;

for each category  $m$  ( $m = 1, 2, \dots, M$ ): generating a category vector  $V_{CAT}(m)$  of order  $N$ , wherein said generating  $V_{CAT}(m)$  comprises for  $n = 1, 2, \dots, N$ : setting  $V_{CAT}(m)[n]$  equal to 1 if category  $m$  has a category key equal to  $CATKEY[n]$ , otherwise setting  $V_{CAT}(m)[n]$  equal to 0;

after said generating  $V_{CAT}(m)$ , for each category  $m$  ( $m = 1, 2, \dots, M$ ): generating a category weight vector  $W_{CAT}(m)$  of order  $N$ , wherein said generating  $W_{CAT}(m)$  comprises for  $n = 1, 2, \dots, N$ : setting  $W_{CAT}(m)[n]$  equal to a second frequency count raised to a power

P<sub>2</sub> greater than 1, wherein the second frequency count consists of a number of appearances, in the collection of stored documents, of the category key associated with V<sub>CAT</sub>(m)[n] if V<sub>CAT</sub>(m)[n] is equal to 1 or consists of 0 if V<sub>CAT</sub>(m)[n] is equal to 0;  
computing distances, wherein said computing distances is selected from the group consisting of computing first distances, computing second distances, computing third distances, and computing fourth distances, wherein said computing first distances comprises computing a dot product of V<sub>DOC</sub> and V<sub>CAT</sub>(m) for m = 1, 2, ..., M, wherein said computing second distances comprises computing a dot product of V<sub>DOC</sub> and W<sub>CAT</sub>(m) for m = 1, 2, ..., M, wherein said computing third distances comprises computing a dot product of W<sub>DOC</sub> and V<sub>CAT</sub>(m) for m = 1, 2, ..., M, and wherein said computing fourth distances comprises computing a dot product of W<sub>DOC</sub> and W<sub>CAT</sub>(m) for m = 1, 2, ..., M;

determining, from categories of a document taxonomy said computed distances, a set of closest categories to the document based on a comparison between the document keys and category keys of the given categories, if said entire document received by said web service host comprises said document keys, or if said web service host has previously performed said generating the plurality of document keys such that said document keys are available to said web service host.

2. (Original) The method of claim 1, further comprising prior to the sending step identifying said web services host, said identifying comprising:

executing a Universal Description, Discovery, and Integration (UDDI) search to identify one or more web services hosts who can receive said document in chunks and who can perform

said at least one of said extracting, generating, and stemming; and  
selecting said web services host from said one or more web services hosts.

3. (Original) The method of claim 1, wherein said transmitting and sequentially transmitting comprise respectively transmitting and sequentially transmitting the first portion and the at least one additional portion via Internet transmission to said web service host.

4. (Currently amended) The method of claim 1, wherein said generating the plurality of document keys comprises:

generating tokens of said text such that stop words do not appear in said tokens; and  
stemming said tokens to generate said document keys from said tokens.

5. (Currently amended) The method of claim 1, wherein ~~said processing comprises said extracting, said generating, and said determining~~ said computing distances consists of said computing first distances.

6. (Currently amended) The method of claim 1, wherein ~~said processing consists of two of said extracting, said generating, and said determining~~ said computing distances consists of said computing second distances.

7. (Currently amended) The method of claim 1, wherein ~~said processing comprises said extracting but not said generating and not said determining~~ said computing distances consists of

said computing third distances.

8. (Currently amended) The method of claim 1, wherein said processing comprises said generating but not said extracting and not said determining said computing distances consists of said computing fourth distances.

9. (Canceled) The method of claim 1, wherein said processing comprises said determining but not said extracting and not said generating.

10. (Currently amended) A system for document analysis and retrieval, comprising a first computing system that includes a remote host, wherein the remote host is remote relative to a web service host in a second computing system, and wherein the remote host is adapted to perform the method of claim 1-

~~transmit a first portion of a document to the web service host, and sequentially transmit at least one additional portion of the document to the web service host, wherein the first portion and the at least one additional portion collectively comprise the entire document, wherein the entire document is adapted to be reconstructed and subsequently processed via processing said entire document by the web service host, said processing comprising at least one of:~~

~~extracting text from said entire document to configure said text in a text format, if said entire document received by said web service host comprises said text in a non-text format; determine~~

generating document keys associated with said text from analysis of said text in said text format, if said entire document received by said web service host comprises said text in said text format, or if said web service host has previously performed said extracting such that said text in said text format is available to said web service host, and determining, from given categories of a document taxonomy, a set of closest categories to the document based on a comparison between the document keys and category keys of the given categories, if said entire document received by said web service host comprises said document keys, or if said web service host has previously performed said generating such that said document keys are available to said web service host.

11-18. (Canceled)

19. (Currently amended) A method for document analysis and retrieval, comprising the following steps of performed in the order recited:

accessing a document taxonomy that comprises M categories such that M is at least 2,  
wherein the document taxonomy is based on a subject matter classification in conjunction with  
a collection of stored documents, wherein each category of the M categories has an associated at  
least one category key, wherein the category keys of all M categories collectively consist of N  
unique category keys sequentially ordered and denoted as CATKEY[1], CATKEY[2], ...,  
CATKEY[N];

receiving, by a web service host in a second computing system from a remote host in a first computing system, a first portion of a document;

sequentially receiving, by the web service host from the remote host, at least one additional portion of the document, wherein the first portion and the at least one additional portion collectively comprise the entire document;

reconstructing the entire document from the first portion and the at least one additional portion; and

processing the entire document by the web service host, wherein said processing comprises at least one of:

extracting text from said entire document to configure said text in a text format, if said entire document received by said web service host comprises said text in a non-text format;

generating a plurality of document keys associated with said text from analysis of said text in said text format, if said entire document received by said web service host comprises said text in said text format, or if said web service host has previously performed said extracting such that said text in said text format is available to said web service host; and

generating a document key vector  $V_{DOC}$  of order  $N$ , wherein said generating  $V_{DOC}$  comprises for  $n=1, 2, \dots, N$ : determining setting  $V_{DOC}[n]$  equal to 1 if the plurality of document keys comprises a document key equal to  $CATKEY[n]$ , otherwise setting  $V_{DOC}[n]$  equal to 0;

after said generating  $V_{DOC}$ , generating a document weight vector  $W_{DOC}$  of order  $N$ ,

wherein said generating  $V_{DOC}$  comprises for  $n = 1, 2, \dots, N$ : setting  $W_{DOC}[n]$  equal to a first frequency count raised to a power  $P_1$  greater than 1, wherein the first frequency count consists of a number of appearances, in the document, of the document key associated with  $V_{DOC}[n]$  if  $V_{DOC}[n]$  is equal to 1 or consists of 0 if  $V_{DOC}[n]$  is equal to 0;

for each category  $m$  ( $m = 1, 2, \dots, M$ ): generating a category vector  $V_{CAT}(m)$  of order  $N$ , wherein said generating  $V_{CAT}(m)$  comprises for  $n = 1, 2, \dots, N$ : setting  $V_{CAT}(m)[n]$  equal to 1 if category  $m$  has a category key equal to equal to  $CATKEY[n]$ , otherwise setting  $V_{CAT}(m)[n]$  equal to 0;

after said generating  $V_{CAT}(m)$ , for each category  $m$  ( $m = 1, 2, \dots, M$ ): generating a category weight vector  $W_{CAT}(m)$  of order  $N$ , wherein said generating  $W_{CAT}(m)$  comprises for  $n = 1, 2, \dots, N$ : setting  $W_{CAT}(m)[n]$  equal to a second frequency count raised to a power  $P_2$  greater than 1, wherein the second frequency count consists of a number of appearances, in the collection of stored documents, of the category key associated with  $V_{CAT}(m)[n]$  if  $V_{CAT}(m)[n]$  is equal to 1 or consists of 0 if  $V_{CAT}(m)[n]$  is equal to 0;

computing distances, wherein said computing distances is selected from the group consisting of computing first distances, computing second distances, computing third distances, and computing fourth distances, wherein said computing first distances comprises computing a dot product of  $V_{DOC}$  and  $V_{CAT}(m)$  for  $m = 1, 2, \dots, M$ , wherein said computing second distances comprises computing a dot product of  $V_{DOC}$  and  $W_{CAT}(m)$  for  $m = 1, 2, \dots, M$ , wherein said computing third distances comprises computing a dot product of  $W_{DOC}$  and  $V_{CAT}(m)$  for  $m = 1, 2, \dots, M$ , and wherein said computing fourth distances comprises computing a dot product of  $W_{DOC}$  and  $W_{CAT}(m)$  for  $m = 1, 2, \dots, M$ ;

determining, from given categories of a document taxonomy said computed distances, a set of closest categories to the document, if said entire document received by said web service host comprises said document keys, or if said web service host has previously performed said generating the plurality of document keys such that said document keys are available to said web service host.

20. (Original) The method of claim 19, wherein the web services host is listed in a Universal Description, Discovery, and Integration (UDDI) registry as being able to receive said document in chunks and being able to perform said at least one of said extracting, generating, and determining.

21. (Original) The method of claim 19, wherein said receiving and sequentially receiving steps comprise receiving the first portion and the at least one additional portion via Internet transmission from said remote host.

22. (Currently amended) The method of claim 19, wherein said generating the plurality of document keys comprises:

generating tokens of said text such that stop words do not appear in said tokens; and stemming said tokens to generate said document keys from said tokens.

23. (Currently amended) The method of claim 19, wherein said processing comprises said extracting, said generating, and said determining said computing distances consists of said

computing first distances.

24. (Currently amended) The method of claim 19, wherein ~~said processing consists of two of said extracting, said generating, and said determining~~ said computing distances consists of said computing second distances.

25. (Currently amended) The method of claim 19, wherein ~~said processing comprises said extracting but not said generating and not said determining~~ said computing distances consists of said computing third distances.

26. (Currently amended) The method of claim 19, wherein ~~said processing comprises said generating but not said extracting and not said determining~~ said computing distances consists of said computing fourth distances

27-28. (Canceled)

29. (Currently amended) The method of claim 19, wherein ~~said processing comprises said determining, and wherein the method further comprises:~~

creating a search string, said search string comprising a logical function of a subset of said document keys;

submitting said search string to a search engine;

receiving links to related documents from said search engine, said links being based on

said search string; and

    returning said links to said remote host.

30. (Currently amended) A system for document analysis and retrieval, comprising a second computing system that includes a web service host, wherein the web service host is remote relative to a remote host in a first computing system, and wherein the web service host is adapted to perform the method of claim 19:

receive a first portion of a document from the remote host;

sequentially receive at least one additional portion of the document from the remote host, wherein the first portion and the at least one additional portion collectively comprise the entire document;

reconstruct the entire document from the first portion and the at least one additional portion; and

implement processing the entire document, said processing comprising at least one of:

extracting text from said entire document to configure said text in a text format, if said entire document received by said web service host comprises said text in a non-text format;

generating document keys associated with said text from analysis of said text in said text format, if said entire document received by said web service host comprises said text in said text format, or if said web service host has previously performed said extracting such that said text in said text format is available to said web service host; and

determining, from given categories of a document taxonomy, a set of closest

categories to the document, if said entire document received by said web service host comprises said document keys, or if said web service host has previously performed said generating such that said document keys are available to said web service host.

31-50. (Canceled)